

Wolf Rainer Less (Hrsg.)
Melanie Fleckenstein
Wolfgang Gottwald
Jens Schröder

Lexikon Qualifizierung analytischer Daten

von der Validierung zur Routine

Vogel Buchverlag

Weitere Informationen:
www.vogel-buchverlag.de

ISBN 978-3-8343-3118-2

1. Auflage. 2011

Alle Rechte, auch der Übersetzung, vorbehalten.
Kein Teil des Werkes darf in irgendeiner Form
(Druck, Fotokopie, Mikrofilm oder einem anderen
Verfahren) ohne schriftliche Genehmigung des
Verlages reproduziert oder unter Verwendung
elektronischer Systeme verarbeitet, vervielfältigt
oder verbreitet werden.

Hier von sind die in §§ 53, 54 UrhG ausdrücklich
genannten Ausnahmefälle nicht berührt.

Printed in Germany

Copyright 2010 by

Vogel Business Media GmbH & Co. KG, Würzburg

Umschlaggrafik:

Vogel Business Media GmbH & Co. KG, Würzburg

Vorwort des Herausgebers

Sehr geehrte Leserinnen und Leser,
sehr geehrte Damen und Herren,

Sie werden über das Buch, das Sie in Ihren Händen halten, sicher etwas erstaunt sein. Ein Lexikon zur Validierung und Qualifizierung analytischer Daten! Nicht wissend, ob Sie eher zu den Analytikern gehören, die in der Thematik «Ergebnisbewertung» erst am Anfang Ihres Engagements stehen, oder ob Sie sich vielleicht doch schon in professionellen Höhen bewegen: Die Form eines Lexikons, die wir für das zu behandelnde Thema gewählt haben, ermöglicht Ihnen eine schnelle und informative Orientierung in allen Bereichen der Bewertung analytischer Daten, ohne umfangreichen Begleittext lesen oder die gewünschte Antwort auf Ihre Frage lange suchen zu müssen.

Manchmal herrscht bei Analytikern Unsicherheit über das zu verwendende «richtige» Verfahren für die jeweils gestellte Validierungsaufgabe, insbesondere dann, wenn eine neue Validierungsstrategie zu entwickeln ist. Die Unsicherheit betrifft z.B. den Detaillierungsgrad, in dem validiert werden muss. Ist die Präzision angemessen? Wann muss gemessen, wann gerechnet werden? Genügt vielleicht schon die Nutzung geeigneter Tabellen, um benötigte Daten zu erhalten? Welche Besonderheiten gelten für die Qualifizierungsstrategien in speziellen Branchen oder in regulierten Prozessen? Das alles sind Fragen, auf die dieses Lexikon jeweils schnelle Einblicke parat hat. Es hilft auch, die richtigen Entscheidungen im geplanten Validierungsaufwand zu treffen. So viel Aufwand wie nötig, so wenig Umstände wie möglich – denn jeder Validierungs- und Bewertungsaufwand verursacht Kosten, und die müssen so niedrig wie möglich gehalten werden.

Wir haben etliche Tabellen im Anhang veröffentlicht, mit deren Hilfe Sie sehr schnell Ihre jeweils benötigten Werte finden können. Das erspart weitgehend das Bemühen anderer Literaturstellen. Zur rechnerischen Lösung von validierungsrelevanten Aufgaben finden Sie die jeweiligen mathematischen Gleichungen oder die dazugehörigen Excel-Befehle. Darüber hinaus ist zur Erläuterung sehr oft eine Beispielsituation eingeschlossen und im direkten Kontext beschrieben, so dass alle Werte und Begriffe verständlich werden. Im Schlussteil des Lexikons wird dann ein umfangreicher Validierungsprozess in größerer Detailtiefe erläutert.

Obwohl die Themen «Validierung» und «Datenqualifizierung» intensiv mit der Mathematik verknüpft sind, haben die Autoren bewusst mathematische Details in den Hintergrund gerückt, dadurch wird die Thematik leichter verständlich. Alle mathematisch Orientierten bitten wir deshalb um Verständnis.

Dieses Lexikon, so unsere Überzeugung, wird Ihnen in allen Fragen der Validierung von großem Nutzen sein.

Frankfurt-Höchst

Wolf Rainer Less
PROVADIS, Partner für Bildung und Beratung

WOLF RAINER LESS,

Jahrgang 1948, absolvierte 1967 seine Ausbildung zum Chemielaboranten. Über seine berufsbegleitenden Weiterbildungen unterschiedlichster Art wurde er 1977, nach seiner Ausbildung zum Labortechniker, Ausbilder für Chemielaboranten im Ausbildungszentrum der Farbwerke Hoechst AG. 1994 übernahm W. R. Less, zu seiner Teamleiterfunktion 1986 in der Produktionstechnik (Berufsgruppe Chemikanten), die Sprecherfunktion der Aus- und Weiterbildung Labortechnik. Seit 1997 ist er Leiter des Kompetenzzenters Labortechnik bei PROVADIS Partner für Bildung und Beratung GmbH, die aus der Hoechst AG hervorgegangen ist. 2001 wurde W. R. Less zusätzlich Geschäftsführer der NOVIA Chromatographie- und Messverfahren GmbH, hatte zwischenzeitlich die Führung des Kompetenzzenters Business Administration bei PROVADIS inne, war Mitglied der AN-Fraktion im Berufsbildungsausschuss der IHK-Frankfurt und Mitglied der Arbeitsgruppe «Neuordnung der Chemielaborantenausbildung» des BAVC Wiesbaden.

B.Sc. MELANIE FLECKENSTEIN,

Jahrgang 1978, absolvierte eine Chemielaborantenausbildung bei der PROVADIS Partner für Bildung und Beratung GmbH in Frankfurt und anschließend eine berufsbegleitende Weiterbildung zur Chemietechnikerin bei der PROVADIS GmbH. Ein berufsbegleitendes Studium zum Bachelor of Science – Fachrichtung Chemical Engineering – wurde 2008 erfolgreich an der PROVADIS School of International Management and Technology abgeschlossen. Nach zunächst zweijähriger Forschungstätigkeit im Pharmaunternehmen Sanofi-Aventis Deutschland GmbH bildete MELANIE FLECKENSTEIN von 2002–2006 Chemielaboranten bei PROVADIS aus. Seit 2007 ist sie in der labortechnischen Weiterbildung der PROVADIS sowie dem Tochterunternehmen NOVIA GmbH mit dem Schwerpunkt Analytische Chemie aktiv. MELANIE FLECKENSTEIN ist Autorin von Fachbüchern für den beruflichen naturwissenschaftlichen Ausbildungsbereich.

B.Sc. JENS SCHRÖDER,

Jahrgang 1979, absolvierte nach einer Ausbildung zum Chemielaboranten bei der PROVADIS International School of Management and Technology ein berufsbegleitendes Studium zum Bachelor of Science, Fachrichtung Chemical Engineering, mit erfolgreichem Abschluss 2007. Von 2002 – 2009 war er Ausbilder für Chemie- und Biologielaboranten bei PROVADIS, neuerdings für Chemikanten und ist darüber hinaus bei der PROVADIS und NOVIA GmbH mit der Erwachsenenweiterbildung betraut, Schwerpunkt Instrumentelle Analytik. JENS SCHRÖDER ist Autor von Fachbüchern für den beruflichen naturwissenschaftlichen Ausbildungsbereich.

Dipl.-Ing. (FH) WOLFGANG GOTTWALD,

Jahrgang 1948, absolvierte nach seiner Ausbildung zum Chemielaboranten bei der Hoechst AG in Frankfurt eine Weiterbildung als Chemietechniker und anschließend, 1983, erfolgreich ein Studium zum Ingenieur für chemische Technologie an der FH Darmstadt. Er war im Ausbildungszentrum der HOECHST AG und der PROVADIS Partner für Bildung und Beratung GmbH jahrelang Ausbilder für Chemielaboranten. Als Bundessachverständiger beteiligte sich WOLFGANG GOTTWALD an der Neuordnung des Chemielaborantenberufs und war über 10 Jahre Mitglied im Fachausschuss für Chemielaboranten bei der PAL in Stuttgart. Im Auftrag für die NOVIA GmbH gestaltet er Weiterbildungs-Seminare im Bereich der Analytischen Chemie und der Qualitätssicherung. WOLFGANG GOTTWALD ist Autor zahlreicher Bücher, die verschiedene Fachbereiche der Analytik und der beruflichen Ausbildung behandeln.

Vorwort der Autoren

«Qualität ist kein Zufall, sie ist immer das Ergebnis angestrengten Nachdenkens!»
JOHN RUSKIN (1819–1900), englischer Sozialökonom

«Die Statistik ist die erste der ungenauen Wissenschaften!»
JULES DE GONCOURT (1830–1870), französischer Sozialromancier

Jeder Analytiker will, dass sein neu entwickeltes analytisches Verfahren bzw. ein in der täglichen Praxis angewandtes Verfahren, richtige und präzise Ergebnisse liefert. Die erhaltenen Analysenergebnisse sollten «gerichtsfest» sein, also einer genauen und intensiven Nachprüfung standhalten. Der Nachweis der Richtigkeit und der Präzision muss dabei so ökonomisch wie möglich durchgeführt werden, denn es handelt sich um einen sehr aufwendigen und teuren Prozess, der gewöhnlich «Validierung» genannt wird. Der Validierungsvorgang selbst ist dabei individuell auf das jeweilige analytische Verfahren zuzuschneiden. Eine allgemeingültige Bewertungsstrategie für angewendete Verfahren gibt es nicht. Dazu kommt, dass der Analytiker mit seinem analytischen Verfahren oft eingebunden ist in eine regulative Umgebung, d.h., dass er unter GMP/GLP oder in einem zertifizierten Bereich arbeitet.

Zur Sicherstellung der Validität analytischer Verfahren werden für die Gewinnung experimenteller Daten meistens allgemeingültige statistische Verfahren ausgewählt. Trotzdem sind viele Anwender unsicher, weil es für sie nicht immer abschätzbar ist, ob die gewählte statistische Rechenprozedur der Intention entspricht. Die Autoren dieses Buches haben in Weiterbildungsseminaren oft die Erfahrung machen müssen, dass diese mit Hilfe von standardisierten Rechenprogrammen viele statistische Kenndaten berechnen und mit den in Spezifikationen erhaltenen Grenzwerten vergleichen, ohne die Bedeutung der Kennwerte richtig gewichten oder in Relation setzen zu können. Aufgrund dessen werden in diesem Buch die bei der branchenübergreifenden Analysenbeurteilung notwendigen Kennwerte und statistischen Tests genau beschrieben. Die dazu nötige Statistik wurde dabei auf das Minimum begrenzt, das unbedingt nötig ist, um die Daten bzw. Ergebnisse korrekt zu interpretieren.

Ein Nachschlagewerk in Form eines Lexikons, das am Arbeitsplatz zur Verfügung steht, schien uns am besten geeignet, Qualitätsprobleme und Fragen bezüglich der Validierung schnell und präzise eruieren zu können. Ein durchgängiges Beispiel zur Validierung eines analytischen Verfahrens am Ende des Buches demonstriert und erläutert beispielhaft die Vorgehensweise.

Dieses Lexikon kann auch in der Ausbildung für den naturwissenschaftlichen Nachwuchs verwendet werden: Auszubildende, Studenten und Diplomanden finden damit zweckdienliche und sichere Analysenmethoden.

Für die Autoren, die alle als Ausbilder und Trainer im analytischen Bereich arbeiten, war es wichtig ein praxisnahes Lexikon vorzulegen, das nicht ausschließlich der «strenge» mathematischen Vorgehensweise folgt, sondern auch der pragmatischen und ökonomischen. So werden z.B. oft in Laboratorien bei der Durchführung von statistischen Tests keine Null- und Alternativhypotesen aufgestellt, sondern Tests pragmatisch mit Schrankenwerten beurteilt. Strenge Mathematiker mögen das nachsehen.

Die hier vorgelegten Daten stammen aus Erfahrungen, der Entwicklung und Umsetzung von Validierungsstrategien von zahlreichen Anwendern aus verschiedenen Branchen, vielen themenbezogenen Diskussionen und unzähligen Fallstudien mit Seminarteilnehmern bei Intensiv-Seminaren der Firma NOVIA GmbH (Leitung Frau Dr. A. Merz, bei der wir uns herzlichst bedanken).

Besonders bedanken wir uns auch bei Herrn Dr. S. Kromidas, der uns vor langer Zeit auf dieses Thema aufmerksam gemacht hat und uns immer mit Rat und Tat zur Seite stand. Auch bei unseren Kollegen von Provadis Partner für Bildung und Beratung GmbH bedanken wir uns für wertvolle Hinweise und Ratschläge.

Über den Onlineservice InfoClick werden beispielhafte Berechnungen in Excel-Tabellen als Vertiefung angeboten. Sie finden diesen kostenlosen Zugangscode auf der ersten Seite des Buchs.

Resonanz zum Buch ist stets willkommen, weil eine lebendige Wissensvermittlung Praxis und Lehrbetrieb immer wieder neu motivieren und inspirieren kann. Den schnellsten Kontakt erfüllt eine E-Mail an: lexikon.validierung@vogel-buchverlag.de.

Frankfurt-Höchst

Melanie Fleckenstein, Wolfgang Gottwald, Jens Schröder

Inhaltsverzeichnis

Vorwort des Herausgebers	5
Vorwort der Autoren	7
Einführung	10
Beispiel einer Validierung	373
Anhang Tabellen	387
Quellenverzeichnis	397
Weiterführende Literatur	401
Übersetzung englischer Fachausdrücke zur Validierung	404

Einführung

Die alphabetische Reihenfolge der nachfolgenden Stichwörter von A bis Z erleichtert Ihnen die gezielte Suche nach dem für Sie relevanten Themengebiet. Sie können zielgerichtet und schnell die passende Antwort auf Ihre Fragen im Laboralltag finden, ohne vorher umfangreiche Theoriekapitel eines Lehrbuches durcharbeiten zu müssen.

Selbstverständlich sind innerhalb der einzelnen Stichwörter Begriffe und Abläufe enthalten, die fachliches Wissen voraussetzen. Um den Umfang des Buches zu begrenzen und um Dopplungen zu vermeiden, sind Begriffe, die von mehreren Stichwörtern erklärt werden, nur einem Stichwort zugeordnet. Es werden dann im Stichwort die jeweils weiterführenden bzw. zum Verständnis gehörenden Zusatzbegriffe als Querverweise mit der typischen grünen Schrift unserer Laborbücher gekennzeichnet.

Bitte berücksichtigen Sie darüber hinaus, dass die in den einzelnen Stichwörtern berechneten Ergebnisse der Beispiele – je nach Vorgehensweise beim Rechnen (Excel/Taschenrechner usw.) und durch Zwischenrundungen in den letzten Nachkommastellen – leicht voneinander abweichen können.

Im Onlineservice **InfoClick** bekommen Sie in Excel ausgearbeitete Beispielberechnungen angeboten. Sie können diesen Service kostenlos mit dem Zugangscode von der 1. Seite des Buchs aufrufen.

Im letzten Teil des Buches finden Sie ein ausführliches exemplarisches Beispiel zur Planung und Durchführung der Validierung einer analytischen Methode. Wir haben Ihnen bei der Beschreibung der Validierung detailliert eine praxisnahe Situation vorgestellt. Darüber hinaus bieten die angegebenen Hinweise der weiterführenden Literatur die Möglichkeit, Zusammenhänge zu vertiefen.

Absoluter Fehler

A

Der absolute Fehler F eines **Messwertes** bzw. eines Ergebnisses definiert sich gemäß Gl. A₁ aus der Differenz des durch die Analyse gefundenen Wertes («Ist-Wert») und dem «richtigen Wert», auch als Soll-Wert bezeichnet.

$$F = \text{Istwert} - \text{Sollwert} \quad (\text{Gl. A}_1)$$

Der absolute Fehler ist mit einem positiven oder negativen Vorzeichen versehen. Die Einheit des absoluten Fehlers ist identisch mit der Einheit des Ist- bzw. Soll-Wertes.

Beispiel

Mittels einer HPLC-Methode wird im Rahmen einer Richtigkeitsüberprüfung eine Wirkstofflösung vermessen, deren Massenkonzentration mit $\beta = 0,15 \text{ g/L}$ genau bekannt ist (Referenz). Der Ist-Wert nach erfolgter Analyse beträgt $\beta = 0,14 \text{ g/L}$. Setzt man die Werte in Gl. A₁ ein, ergibt sich gemäß Gl. A₂ ein absoluter Fehler von $F = -0,01 \text{ g/L}$.

$$F = 0,14 \text{ g/L} - 0,15 \text{ g/L} = -0,01 \text{ g/L} \quad (\text{Gl. A}_2)$$

Der relative **Fehler f** ergibt sich durch Bildung des Quotienten aus absolutem Fehler und bekanntem Soll-Wert.

Abweichung

Als Abweichung wird bei einer Messung die Differenz zwischen dem **Messwert** und dem **Erwartungswert** (wahrer Wert) bezeichnet.

Dies kann z.B. die Differenz der Anzeige eines UV-Vis-Spektralfotometers sein, das für einen Filter eine Extinktion von 0,588 anzeigt, wobei der Erwartungswert bei 0,595 liegt. Die Abweichung beträgt hier 0,007. Zeigt das Fotometer diese Abweichung über mehrere Prüfperioden an, spricht man von einer systematischen Abweichung (**systematischer Fehler**). Wird nur bei einer Prüfperiode dieses Messergebnis erhalten, nennt man dies zufällige Abweichung (**zufälliger Fehler**). Jegliche Abweichungen sollten auf ihre Ursache hin untersucht werden.

Statistische Größen wie die **Standardabweichung s** , **Reststandardabweichung s_y** oder die relative **Verfahrensstandardabweichung V_{x_0}** werden zur Bewertung von Messwerten und deren Abweichung berechnet.

Der Begriff Abweichung wird auch im **Qualitätsmanagement** verwendet. Eine Abweichung ist in diesem Bereich z.B. die Nichteinhaltung einer bestehenden **Standardarbeitsanweisung**. Im Laborumfeld wird häufig anstatt von Abweichung der Begriff **Fehler** verwendet. Auch der englische Begriff **Bias** beschreibt die Abweichung von Ist-Wert und dem wahren Wert.

Accuracy Profile

Im Allgemeinen werden zur Beurteilung von Validierungsdaten diese tabellarisch zusammengetragen und dann bewertet. Bezüglich der Formatierung und Aufbereitung gibt es hier natürlich enorme Unterschiede in der Übersichtlichkeit und Verständlichkeit dieser individuellen Datenzusammenstellungen. Eine Kommission der SFSTP (Société Française des Sciences et Techniques Pharmaceutiques) empfiehlt als Harmo-

nisierungsansatz die grafische Darstellung der Validierungsdaten, das sog. «Accuracy Profile».

Der prinzipielle Aufbau eines «Accuracy Profiles» kann Bild A₁ entnommen werden.

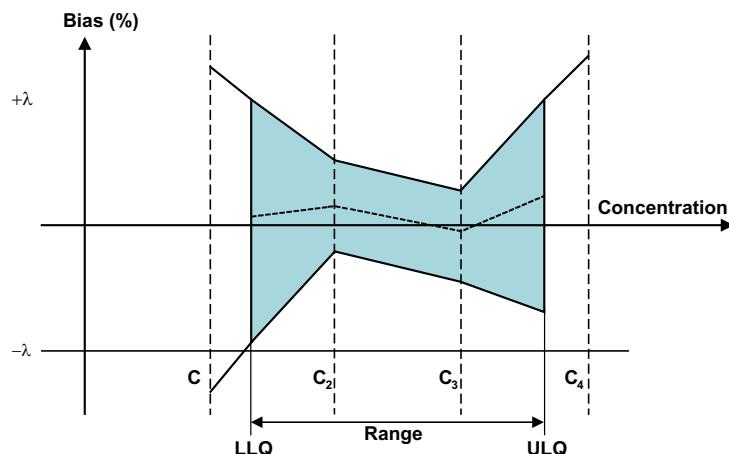


Bild A₁ Prinzipieller Aufbau eines «Accuracy Profiles»

Auf der Abszisse ist die Konzentration c aufgetragen, auf der Ordinate der **systatische Fehler (Bias)**. Die Bezeichnungen LLQ und ULQ stehen für die untere und obere Grenze des **Arbeitsbereichs**. Der Bereich von $-\lambda \dots +\lambda$ wird durch die Spezifikation vorgegeben.

Die **Richtigkeit** über die **Wiederfindungsrate WFR** wurde auf den Konzentrationsniveaus C₁...C₄ bestimmt, die gestrichelte Linie gibt den mittleren Verlauf wieder. Man erkennt, dass die WFR in Abhängigkeit von der Konzentration c keinem **Trend** unterliegt. Die graue Fläche ergibt sich durch Verbinden der Ergebnisse der Messungen des **zufälligen Fehlers (Vertrauensbereich)**. Man erkennt aus Bild A₁, dass die Streuungen der Ergebnisse an den Rändern des Arbeitsbereiches größer sind.

Achsenabschnitt a

Der Achsenabschnitt a (Ordinatenabschnitt) entspricht dem Abschnitt, der entsteht, wenn eine Funktion die y -Achse (also bei $x = 0$) schneidet (Bild A₂). Bei einer linearen **Kalibrierfunktion** nach Gl. A₃ entspricht der Term a dem Achsenabschnitt:

$$y = b \cdot x + a \quad (\text{Gl. A}_3)$$

mit:

- y abhängige Variable (Messgröße)
- b Steigung (Empfindlichkeit)
- x unabhängige Variable (z.B. Konzentration c)
- a Achsenabschnitt

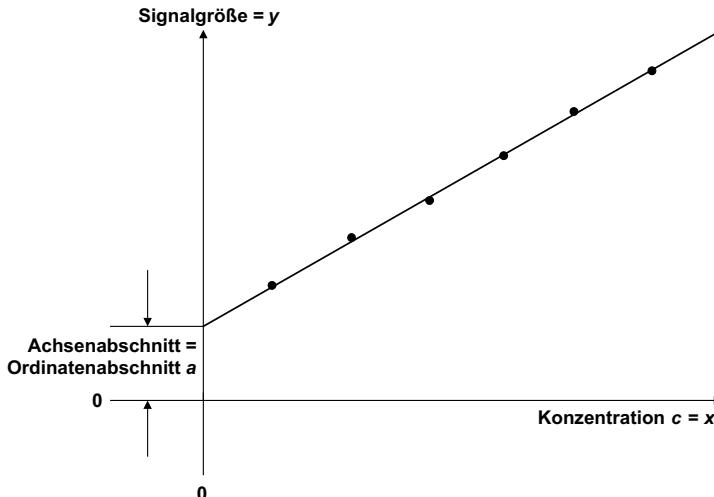


Bild A₂ Lineare Funktion und Achsenabschnitt a

Bei einer Polynomfunktion 2. Grades nach Gl. A₄ entspricht a ebenfalls dem Achsenabschnitt.

$$y = c \cdot x^2 + b \cdot x + a \quad (\text{Gl. A}_4)$$

Formal gesehen, ist der Achsenabschnitt a der **Leerwert (Blindwert)** der analytischen Methode, da hier definitionsgemäß die Konzentration $c = x = 0$ ist. Das gilt jedoch nur, wenn die Kalibrierung in der Nähe der Konzentration $c = 0$ erfolgte bzw. wenn **Varianzenhomogenität** über den ganzen **Arbeitsbereich** herrscht.

Die Koeffizienten b , a (und c) werden mit Hilfe von mathematischen **Regressions** aus den Analysesignalen y_i und den dazugehörigen Konzentrationen $c = x_i$ erhalten. Ist der Achsenabschnitt a einer Analysenfunktion **signifikant** von 0 verschieden, so enthält die Funktion einen systematischen **Fehler**. Die Prüfung, ob der Achsenabschnitt signifikant von $a = 0$ verschieden ist, kann durch die Berechnung des **Prognosebereichs** bei $x = c = 0$ mit $P = 95\%$ erfolgen. Schließt der Prognosebereich des Achsenabschnitts den «0»-Wert mit ein, so ist zwar formal ein Achsenabschnitt vorhanden, aber er ist nicht signifikant von «0» verschieden und kann statistisch vernachlässigt werden (Bild A₃).

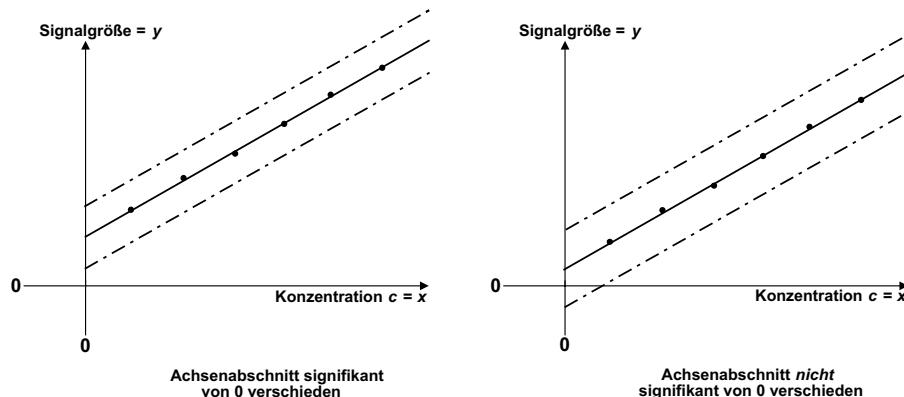


Bild A₃ Achsenabschnitt signifikant und nicht signifikant von «0» verschieden

Schließt der Prognosebereich den «0»-Wert mit ein, ist eine lineare Funktion auch «proportional». Mit diesem Analysenverfahren kann u.U. eine **1-Punkt-Kalibrierung** durchgeführt werden.

Ist der Achsenabschnitt signifikant größer oder kleiner als $a = 0$, kann der systematische **Fehler** dadurch eliminiert werden, dass beim Auswertevorgang der Achsenabschnitt a mit in die Rechnung eingeht.

Ein negativer Achsenabschnitt a , bei dem die Funktion die y -Achse unterhalb von $y = 0$ schneidet, weist meistens auf eine nicht ausreichend vorhandene Selektivität des Verfahrens hin.

Akkreditierung

Unter einer Akkreditierung versteht man allgemein den Nachweis und die Anerkennung der Kompetenz, mit der ein bestimmter Auftrag, eine genau definierte Prüfung oder eine bestimmte Dienstleistung ausgeführt wird. Hauptziel einer Akkreditierung ist der Abbau von Handelsbeschränkungen, die z.B. durch Mehrfachanalysen beim Produzenten und beim Käufer bzw. Kunden entstehen können. Durch die gegenseitige Anerkennung vergleichbarer Ergebnisse werden diese Beschränkungen und Mehrfachanalysen abgebaut oder ganz beseitigt. Das schließt ein, dass das akkreditierte Labor jederzeit richtige und präzise Ergebnisse liefert und unabhängig vom Einfluss des Auftraggebers handelt.

Die Akkreditierung für die Arbeitsweise von Prüf- und Kalibrierungslaboratorien wird für den Laborbereich über die **DIN EN ISO 17 025** geregelt. Das Labor wird von einer Expertengruppe einer unabhängigen Akkreditierungsstelle mittels Begehungen sowie Audits (**Auditierung**) geprüft und nach der Erteilung der Akkreditierung in jährlichen Begehungen überwacht. Die überwachenden Organisationen sind sog. Akkreditierungsstellen, wie z.B. DACH und DAP. Die ISO 17 025 enthält u.a. [A1]:

- Führungs- und Managementanforderungen an ein wirksames QS-System,
- Arbeitsweisen zur Durchführung kompetenter Analysen,
- Kalibrierung** und Qualifizierung von Geräten,
- Validierung** von Messverfahren,
- Erstellungsanforderungen von Prüfberichten und
- Interpretation von Ergebnissen.

Näheres siehe **DIN EN ISO 17 025**. Die Kompetenz des Prüflabors kann auch über die Beteiligung an Ringversuchen beurteilt werden.

Akzeptanzkriterium

Unter einem Akzeptanzkriterium (lat. «accipere» für annehmen, billigen, gutheißen) versteht man ein schlüssiges Kriterium zur Annahme oder zur Ablehnung eines Ergebnisses. Annahme oder Zurückweisung des Ergebnisses entscheiden über die jeweilige zu belegende Aussage (z.B. über die Präzision).

Beispiel

Wird der Akzeptanzwert bei der Präzisionsuntersuchung auf den **Variationskoeffizienten VK** $< 2\%$ gelegt und man erhält aus einem **Wiederholexperiment** einen $VK = 3,5\%$, dann ist das Akzeptanzkriterium nicht erfüllt und die Präzision des analytischen Verfahrens reicht nicht aus. Die Präzision muss durch geeignete Maßnahmen verbessert werden. Wäre $VK < 2\%$, sollte trotzdem überprüft werden, ob keine Ausreißer und keine Trends in der Datenreihe enthalten sind.

Alpha-Fehler (α -Fehler)

Diesen Fehler nennt man auch «Fehler 1. Art», «Produzentenrisiko» oder «falscher Alarm».

Bei jedem statistischen Test wird zunächst eine «Nullhypothese» H_0 aufgestellt. Darunter versteht man eine Behauptung, die eine grundlegende Aussage über die **Grundgesamtheit** macht. Die 2. Behauptung, die im Gegensatz zur Nullhypothese H_0 steht, nennt man «Alternativhypothese» H_1 (oder Gegenhypothese).

Bei der Auswahl, ob die Nullhypothese H_0 nicht abgelehnt («nicht nein») oder abgelehnt («nein») wird, können natürlich Fehler entstehen. Je nach dem, wie sich der gemachte Fehler auswirkt, unterscheidet man in Alpha-Fehler (Fehler 1. Art) und **Beta-Fehler** (Fehler 2. Art). Die Einteilung der Fehler wird am Beispiel eines **Ausreißer-Tests** erläutert.

Als Nullhypothese H_0 wird die Aussage «die Datenreihe ist ausreißerfrei» aufgestellt, die Alternativhypothese H_1 ist demnach «die Datenreihe enthält Ausreißer». Wird die Nullhypothese H_0 «die Datenreihe ist ausreißerfrei» aufgrund eines Tests irrtümlicherweise abgelehnt, obwohl sie richtig gewesen wäre (sie enthält tatsächlich keinen Ausreißer!), nennt man diesen Fehler «Alpha-Fehler» oder «Fehler 1. Art». Das Auftreten des Alpha-Fehlers ist unvermeidlich, man kann aber dafür sorgen, dass die Nullhypothese H_0 nicht voreilig abgelehnt wird.

Der β -Fehler oder «Fehler 2. Art» ist der Fehler, der entsteht, wenn die Nullhypothese H_0 «die Datenreihe ist ausreißerfrei» nicht abgelehnt wird, obwohl sie falsch ist, weil sie Ausreißer enthält. In Tabelle A₁ sind beide Fehlerarten zusammengefasst.

Tabelle A₁ Hypothese und Fehlerarten

	Nullhypothese H_0 wird nicht abgelehnt	Nullhypothese H_0 wird abgelehnt
Nullhypothese H_0 ist wahr	richtige Entscheidung	Alpha-Fehler (1. Art)
Nullhypothese H_0 ist falsch	Beta-Fehler (2. Art)	richtige Entscheidung

Alternativhypothese H_1

Die Behauptung, die eine grundlegende Aussage über die **Grundgesamtheit** macht, bezeichnet man als Nullhypothese H_0 . Die 2. Behauptung, die im Gegensatz zur Nullhypothese H_0 steht, nennt man «Alternativhypothese» H_1 (oder Gegenhypothese). Mit einem **statistischen Test** kann man mit einer vorgegebenen Wahrscheinlichkeit (bzw. Irrtumswahrscheinlichkeit) feststellen, ob die Nullhypothese H_0 nicht abgelehnt («nicht nein») oder abgelehnt («nein») werden soll. Im letzten Fall wird der Alternativhypothese H_1 der Vorzug gegeben. Beispiele findet man unter **Nullhypothese H_0** .

Analysenergebnis

Das Analysenergebnis ist eine berechnete Größe aus Messwerten. Eine Messung an einem Gerät ergibt einen **Messwert**. Dieser Messwert kann z.B. die Extinktion einer Coffeinlösung am UV-Vis-Spektralphotometer sein. Durch eine vorher gefundene, gültige **Kalibrierfunktion** kann aus dem Messwert die Konzentration an Coffein berechnet werden. Jede Messung unterliegt einer gewissen **Messunsicherheit**. Nach einer Mehrfachbestimmung der Extinktion, z.B. der Coffeinlösung, wird aus den berechneten Analysenergebnissen der **Mittelwert** \bar{x} gebildet.

Dieser Mittelwert wird häufig mit dem berechneten **Vertrauensbereich VB** zusammen als Analysenergebnis angegeben, z.B.:
 $m(\text{Coffein}) = \bar{x} \pm VB = 0,2 \text{ mg/L} \pm 0,003 \text{ mg/L}$.

Analysenmethode

In der Analytik gibt es verschiedene Analysenmethoden. Dazu zählen z.B. HPLC, GC, UV-Vis, DC, MS und die Titration. Muss für eine neue Aufgabenstellung eine geeignete Analysenmethode ausgesucht werden, sind verschiedene Aspekte zu beachten.

- Welche Methode kann aufgrund der physikalischen Beschaffenheit der Probe eingesetzt werden?
- Ist die Methode selektiv genug für meine Problemstellung?
- Können geforderte **Spezifikationen** (z.B. **Standardabweichung**) eingehalten werden?

Weitere Randbedingungen wie Personalbedarf und zeitlicher Aufwand sind ebenso zu berücksichtigen. Im Pharmabereich, im Pflanzenschutz und in der Lebensmittelindustrie sind Analysenmethoden und die **Analysenverfahren** häufig vorgeschrieben (z.B. **Arzneibuch**).

Analysenprobe

Die Analysenprobe ist die Probe, die direkt mit dem **Analysenverfahren** untersucht wird. Sie kann durch verschiedene Arten der Aufarbeitung wie z.B. Verdünnen, Filtrieren, Säulenchromatographie oder Extraktion aus der **Urprobe** gewonnen werden. Bei der Aufarbeitung soll der **Analyt** in der Analysenprobe nicht verändert werden. Ausnahmen sind gezielte Derivatisierungen (z.B. Austausch einer polaren Gruppe durch eine weniger polare Methylgruppe zur Erhöhung der thermischen Stabilität für eine GC-Analyse). In seltenen Fällen kann die Urprobe direkt als Analysenprobe untersucht werden.

Analysenverfahren

Das Analysenverfahren umfasst alle Tätigkeiten rund um eine Analyse. Dies beginnt bei der Auswahl einer geeigneten **Analysenmethode**. Zum Verfahren gehören weiterhin die Vorbereitung der Probenahme und die Probenahme selbst. Es folgt die Aufarbeitung der Probe (Probenvorbereitung), die eigentliche Messung der Probe und die Auswertung. Ergänzend schließt die Dokumentation aller vorangegangenen Teilschritte das Analysenverfahren ab. In vielen Anwendungsbereichen dürfen nur Verfahren eingesetzt werden, deren Eignung für die Aufgabenstellung zuvor durch eine **Validierung** bewiesen wurde. Im Bereich der Pharma- und Lebensmittelindustrie gibt es eine Vielzahl an vorgeschriebenen Analysenverfahren (z.B. im **Arzneibuch**).

ANOVA, einfaktorielle

Mit einer einfachen oder einfaktoriellen Varianzanalyse (*Analysis of Variances, ANOVA*) kann geprüft werden, ob und wie sich eine Datenserie (Reihe) von einer anderen Serie unterscheidet. Dazu werden **Streuungen innerhalb** und **zwischen** Serien statistisch verglichen. ANOVA ist die mathematische Grundlage zur Berechnung von

Daten aus **Ringversuchen**. Kenndaten, die mit Hilfe der ANOVA erhalten werden, können zur Bewertung der **Robustheit** dienen. Die Vorgehensweise bei der Durchführung der ANOVA veranschaulicht ein Beispiel.

Beispiel

Für einen ringversuchähnlichen Vergleich mehrerer Laboratorien soll eine stabile und homogene Probe jeweils 4-mal inklusive aller Analysenschritte analysiert werden. Die Werte sollen mit Hilfe der ANOVA daraufhin überprüft werden, ob die Ergebnisse **zwischen** den Laboratorien signifikant oder nur zufällig größere Streuungen aufweisen als die Werte, die jeweils **innerhalb** der Laboratorien ermittelt wurden. Ist ein signifikanter Unterschied zwischen den beiden Streuungsarten («zwischen» und «innerhalb») nachweisbar, weist dies oft auf ein nicht ausreichend robustes analytisches Verfahren hin.

Es wurden in 4 Laboratorien mit jeweils der gleichen Probe je 4 Bestimmungen vorgenommen. Die Ergebnisse sind in Tabelle A₂ aufgeführt. Zugunsten der Übersicht sind die Einheiten nicht aufgeführt.

Tabelle A₂ Ergebnisse der Untersuchung (ppm Analyt)

	Labor Nr. 1 $j = 1$	Labor Nr. 2 $j = 2$	Labor Nr. 3 $j = 3$	Labor Nr. 4 $j = 4$
Probe Nr. 1, $i = 1$	231	244	225	237
Probe Nr. 2, $i = 2$	238	235	234	237
Probe Nr. 3, $i = 3$	231	241	231	231
Probe Nr. 4, $i = 4$	229	239	229	242

Zunächst werden für die Werte innerhalb jedes Laboratoriums (von $j = 1$ bis $j = 4$) der **Mittelwert \bar{x}** und die **Standardabweichung s** berechnet (Tabelle A₃).

Tabelle A₃ Mittelwerte und Standardabweichungen (ppm Analyt)

	Labor Nr. 1 $j = 1$	Labor Nr. 2 $j = 2$	Labor Nr. 3 $j = 3$	Labor Nr. 4 $j = 4$
Probe Nr. 1, $i = 1$	231	244	225	237
Probe Nr. 2, $i = 2$	238	235	234	237
Probe ,Nr. 3, $i = 3$	231	241	231	231
Probe Nr. 4, $i = 4$	229	239	229	242
Anzahl L_j	4	4	4	4
Mittelwert \bar{x}_j	232,25	239,75	229,75	236,75
Standard-abweichung s_j	3,95	3,77	3,77	4,50

Bei allen Kenndaten und den daraus abgeleiteten Größen wird vorausgesetzt, dass die Einzelwerte nur «zufällig» vom Mittelwert abweichen. Daher sollte man entweder die Mittelwerte mit einem Test der **«MANDEL-h-Statistik»** oder mit einem **Ausreißertest** vor der eigentlichen ANOVA überprüfen (COCHRAN-Test mit $P = 95\%$; F-Test mit $P = 99\%$).

Ein Ausreißer-Test nach **GRUBBS** auf dem Niveau von $P = 95\%$ ergab, dass keiner der 4 Mittelwerte als **Ausreißer** zu erkennen ist.

Als nächstes muss überprüft werden, ob alle Einzelstandardabweichungen $s_1 \dots s_4$ aus «gleichwertigen» Datenreihen stammen, d.h., ob sie einer gemeinsamen Grundgesamtheit angehören können. Bei 2 Datenreihen wäre der **F-Test** anwendbar. Liegen aber mehr als 2 Datenreihen (Serien) vor und ist die Datenanzahl in jeder Serie gleich, kann zur Überprüfung der **Varianzenhomogenität** der **COCHRAN-Test** benutzt werden. (Die Einheiten wurden zugunsten der Übersichtlichkeit weggelassen.)

In Tabelle A₃ enthält Labor Nr. 4 mit $s_4 = 4,50$ die größte Standardabweichung, daher wird die Varianz s^2 dieses Laboratoriums Nr. 4 im Zähler der Gl. A₅ berücksichtigt.

$$PG = \frac{s_{\max}^2}{\sum s_j^2} = \frac{4,50^2}{3,95^2 + 3,77^2 + 3,77^2 + 4,50^2} = \underline{0,315} \quad (\text{Gl. A}_5)$$

Die Prüfgröße PG wird mit der **COCHRAN-Tabelle** (siehe Anhang) verglichen, wobei k die Anzahl der Laboratorien ist ($k = 4$) und f die Anzahl der **Freiheitsgrade** pro Labor ($f = N_j - 1 = 4 - 1 = 3$). Aus der Tabelle mit $P = 95\%$ ($\alpha = 0,05$) kann der Schrankenwert 0,6841 entnommen werden. Da die Prüfgröße PG kleiner als der Schrankenwert (Tabellenwert) ist, kann nach **COCHRAN** auf dem Niveau von $P = 95\%$ davon ausgegangen werden, dass Varianzenhomogenität zwischen allen Laboratorien herrscht.

Als nächstes wird die «durchschnittliche Standardabweichung» s_r aus allen Laborstandardabweichungen s_j berechnet. Es gelten Gl. A₆ und Gl. A₇:

$$N = \sum_{j=1}^k L_j = 4 + 4 + 4 + 4 = \underline{16} \quad (\text{Gl. A}_6)$$

$$s_r^2 = \frac{1}{N-k} \cdot \sum_{j=1}^k (L_j - 1) \cdot s_j^2 \quad (\text{Gl. A}_7)$$

mit:

- N Gesamtzahl aller Analysen ($N = 16$)
- k Gesamtzahl der Laboratorien ($k = 4$)
- L_j Analysen pro Laboratorium ($L_1 \dots L_4 = 4$)
- s_j Standardabweichung der Laboratorien $j = 1$ bis $j = 4$
- s_r^2 mittlere Varianz «innerhalb der Laboratorien»

Die mittlere Varianz s_r^2 innerhalb aller Laboratorien (Serien) beträgt nach Gl. A₈:

$$s_r^2 = \frac{1}{16-4} \cdot ((4-1) \cdot 3,95^2 + (4-1) \cdot 3,77^2 + (4-1) \cdot 3,77^2 + (4-1) \cdot 4,50^2) = \underline{16,07} \quad (\text{Gl. A}_8)$$

Die daraus durch Radizieren berechnete Standardabweichung s_r repräsentiert eine durchschnittliche Standardabweichung **innerhalb** der Laboratorien (Gl. A₉):

$$s_r = \sqrt{s_r^2} = \sqrt{16,07} = \underline{4,01} \quad (\text{Gl. A}_9)$$

Die Standardabweichung s_r wird auch als **Wiederholstandardabweichung** bezeichnet.

Als nächstes wird die Streuung **zwischen** den 4 Laboratorien berechnet. Dazu muss zunächst ein Gesamtmittelwert \bar{x} (das ist der «Mittelwert aller Mittelwerte») nach Gl. A₁₀ berechnet werden.

$$\hat{x} = \frac{1}{N} \cdot \sum_{j=1}^k L_j \cdot \bar{x}_j \quad (\text{Gl. A}_{10})$$

Für die vorliegende Messreihe kann nach Gl. A₁₁ ein Gesamtmittelwert \hat{x} von 234,63 berechnet werden.

$$\hat{x} = \frac{1}{16} \cdot (4 \cdot 232,25 + 4 \cdot 239,75 + 4 \cdot 229,75 + 4 \cdot 236,75) = \underline{\underline{234,63}} \quad (\text{Gl. A}_{11})$$

Die Varianz der Werte **zwischen** den Laboratorien (s_z^2) kann mit Gl. A₁₂ berechnet werden.

$$s_z^2 = \frac{1}{k-1} \cdot \sum_{j=1}^k L_j \cdot (\bar{x}_j - \hat{x})^2 \quad (\text{Gl. A}_{12})$$

$$s_z^2 = \frac{1}{4-1} \cdot [4 \cdot (232,25 - 234,63)^2 + \dots + 4 \cdot (236,75 - 234,63)^2] = \underline{\underline{80,25}} \quad (\text{Gl. A}_{13})$$

Durch Radizieren des Ergebnisses aus Gl. A₁₃ erhält man die Standardabweichung der Daten **zwischen** den Laboratorien (Serien):

$$s_z = \sqrt{80,25} = \underline{\underline{8,96}} \text{ ppm} \quad (\text{Gl. A}_{14})$$

Beim Vergleich der beiden Standardabweichungen («innerhalb der Laboratorien» und «zwischen den Laboratorien») fällt auf, dass der Wert «zwischen den Laboratorien» deutlich größer ist, als der «in den Laboratorien». Es muss nun geprüft werden, ob beide Werte «signifikant» oder nur «zufällig» voneinander verschieden sind. Eine Überprüfung, ob der Unterschied «signifikant» oder nur «zufällig» ist, kann mit dem **F-Test** vorgenommen werden. Es wird das Signifikanzniveau von $P = 99\%$ empfohlen. Zur Berechnung der Prüfgröße PG wird die größere Varianz durch die kleinere dividiert (Gl. A₁₅):

$$PG = \frac{s_z^2}{s_r^2} = \frac{s_z^2}{s_r^2} = \frac{80,25}{16,07} = \underline{\underline{4,99}} \quad (\text{Gl. A}_{15})$$

Mit $f_1 = k - 1 = 3$ und $f_2 = N - k = 12$ kann aus der F-Tabelle (im Anhang Anh₄) ein Schrankenwert von 5,953 ($P = 99\%$) abgelesen werden. Auf diesem Niveau kann ein signifikanter Unterschied nicht nachgewiesen werden ($4,99 < 5,953$), die Unterschiede sind zufällig. Die Ergebnisse weisen darauf hin, dass das Verfahren ausreichend robust sein könnte.

Sollte bei einer ANOVA-Untersuchung («zwischen den Laboratorien») s_z kleiner sein als s_r («innerhalb der Laboratorien»), geht man davon aus, dass keine signifikanten Abweichungen bestehen und der F-Test ganz entfallen kann.

Die **Vergleichsstandardabweichung** s_R setzt sich aus der Wiederholstandardabweichung s_r und der Standardabweichung zwischen den Laboratorien s_z zusammen (Gl. A₁₆).

$$s_R = \sqrt{s_r^2 + s_z^2} \quad (\text{Gl. A}_{16})$$

Für unser Beispiel wäre die aus der Wiederholstandardabweichung und der Standardabweichung zwischen den Laboratorien zusammengesetzte Vergleichsstandardabweichung s_R nach Gl. A₁₇ $s_R = 9,81$.

$$s_R = \sqrt{16,07 + 80,25} = \underline{\underline{9,81}} \quad (\text{Gl. A}_{17})$$

Die einfaktorielle ANOVA ist die Grundlage zur Berechnung von Ringversuchsdaten. Allerdings sollten bei [Ringversuchen](#) nach DIN 38 402 in mindestens $k = 8$ Laboratorien die Mindestanzahl von $M = 3$ Proben untersucht werden. Die 3 Proben sollten den gesamten Arbeitsbereich abdecken. Dabei wird eine $L = 4$ -fach-Bestimmung pro Probe empfohlen.

Eine «einfaktorielle ANOVA» kann auch mit [Excel-Funktionen](#) durchgeführt werden. Dazu kann der Funktionsassistent benutzt werden: Unter Hauptmenü/Extras können die «Funktionsanalysen» aufgerufen werden. Ggf. sind sie vorher zu aktivieren (siehe Excel-Handbuch).

Nach der Auswahl der «einfaktoriellen Varianzanalyse» kann «alpha» (hier 0,01 für $P = 99\%$) eingegeben und der Tabellenbereich (z.B. neues Arbeitsblatt) benannt werden (geordnet nach Spalten). Man erhält die Daten nach Tabelle A₄.

Tabelle A₄ Ergebnis der Excel-ANOVA

ANOVA: einfaktorielle Varianzenanalyse

Zusammenfassung

Gruppen	Anzahl	Summe	Mittelwert	Varianz
Spalte 1	4	925	232,25	15,5833
Spalte 2	4	959	239,75	14,25
Spalte 3	4	919	229,75	14,25
Spalte 4	4	947	236,75	20,25

ANOVA

Streuungs-ursache	Quadrat-summen (SS)	Freiheits-grade (df)	mittlere Quadrat-summe (MS)	Prüfgröße (F)	kritischer F-Wert
Unterschiede zwischen den Gruppen	240,75	3	80,25	4,9896	5,95252914
innerhalb der Gruppen	103	12	16,083		
Gesamt	433,75	15			

Man erkennt, dass in der Tabelle die Prüfgröße PG (4,9896), die beiden Varianzen s^2_r und (80,25 und 16,08) und der kritische F -Wert (5,952...) mit den manuell berechneten Werten nahezu identisch sind.

Äquivalenz-Test

Durch einen statistischen Test ist es möglich, einen Unterschied zweier Parameter (z.B. zwischen einem **Mittelwert** \bar{x} und einem **Sollwert** S) auf einem bestimmten Niveau nachzuweisen. In der Praxis müssen aber oft Unterschiede zu einer «starren» kundenorientierten Spezifikationsgrenze nachgewiesen werden, die vom Kunden festgelegt und damit nicht von statistischen Kenngrößen abhängig ist. Diese Unterschiedsfeststellung leistet ein Äquivalenztest. 2 Äquivalenz-Tests werden hier vorgestellt:

- Äquivalenztest 1: Unterschied Mittelwert zu Sollwert,
- Äquivalenztest 2: Unterschied zweier Mittelwerte.

1. Äquivalenztest: Unterschied Mittelwert zu Sollwert

Der Unterschied zwischen dem Mittelwert \bar{x} von Analysenergebnissen einer Stichprobenreihe und einem vorgegebenen Sollwert S wird als Maß für den systematischen **Fehler** und damit für die Richtigkeit herangezogen. Dieser Unterschied wird bei einem **Sollwert-t-Test** oft aufgrund kleiner Standardabweichungen (also durch eine hohe Präzision) signifikant, obwohl der Unterschied zwischen dem Sollwert S und dem Mittelwert \bar{x} für den Kunden durchaus noch akzeptabel wäre. In Gl. A₁₈ zur Berechnung der Prüfgröße PG für den Sollwert-t-Test lässt sich der Einfluss der Standardabweichung s auf die Prüfgröße PG gut erkennen:

$$PG = \frac{|\bar{x} - S|}{s} \cdot \sqrt{N} \quad (\text{Gl. A}_{18})$$

mit:

- PG Prüfgröße
- \bar{x} Mittelwert
- s Standardabweichung
- S Sollwert
- N Anzahl der Proben

Bei einer kleinen Standardabweichung s (hohe Präzision) wird die Prüfgröße PG groß und überschreitet dann oft den zulässigen Schrankenwert.

Was jedoch vielfach in der Praxis benötigt wird, ist eine Untersuchung, ob der Unterschied zwischen dem Sollwert S und dem Mittelwert \bar{x} für den Kunden **effektiv** «signifikant» oder «nicht signifikant» ist. Dazu sind kundenspezifische Spezifikationsgrenzen notwendig.

Vorgegeben ist eine **Referenzlösung** mit einem Sollwert S und als Spezifikationsgrenze ein für die Praxis akzeptabler Abstand A (z.B. $\pm 2\%$). Weiterhin wird mit einem zu prüfenden Analysenverfahren die Referenzlösung unter **Wiederholbedingungen** mehrfach analysiert. Man erhält den **Mittelwert** \bar{x} der Stichprobenreihe und die **Standardabweichung** s des Verfahrens. Daraus lässt sich nach Gl. A₁₉ das **Konfidenzintervall** KI berechnen:

$$KI = \bar{x} \pm \frac{t \cdot s}{\sqrt{N}} \quad (\text{Gl. A}_{19})$$

Es könnte nun sein, dass eine der beiden Grenzen des Konfidenzintervalls den akzeptablen Abstand des Sollwertes nicht mehr mit einschließt (Bild A₄). Dies würde ein klassischer Sollwert-t-Test übersehen, denn der Sollwert-t-Test prüft nur den Abstand des Sollwertes S zum Mittelwert \bar{x} .

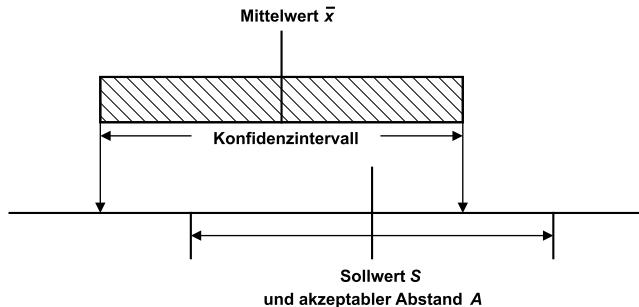


Bild A₄ Messwert, Konfidenzintervall und akzeptabler Abstand A des Sollwertes

Der Äquivalenz-Test unterscheidet sich vom klassischen Sollwert- t -Test durch einen Vergleich zweier Intervalle. Das 1. Intervall entsteht, wenn der Sollwert S einer unteren bzw. oberen Akzeptanzgrenze A zugeordnet wird. Diese Akzeptanzgrenzen können durch Behörden, Regularien oder durch firmeninterne Vorgaben bestimmt sein. Im Rahmen analytischer Transferaufgaben wird z.B. oft eine Akzeptanzgrenze von $A = \pm 2\%$ vorgeschlagen (Bild A₅).

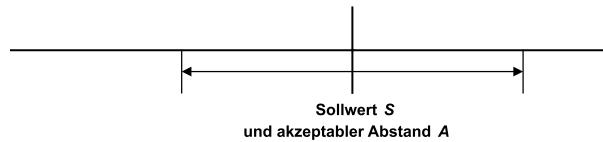


Bild A₅ Sollwert S und akzeptabler Abstand A

Das 2. Intervall wird durch die obere bzw. untere Grenze des Konfidenzintervalls nach Gl. A₂₀ und Gl. A₂₁ gebildet. Das Testen der Intervalle mit Hilfe des Äquivalenztests erfolgt über den Vergleich des Konfidenzintervalls mit den vorher festgelegten Akzeptanzgrenzen. Sobald eine der Grenzen des Konfidenzintervalls **außerhalb** des Akzeptanzbereiches liegt, wird auf **Nichtäquivalenz** geschlossen (Bild A₆).

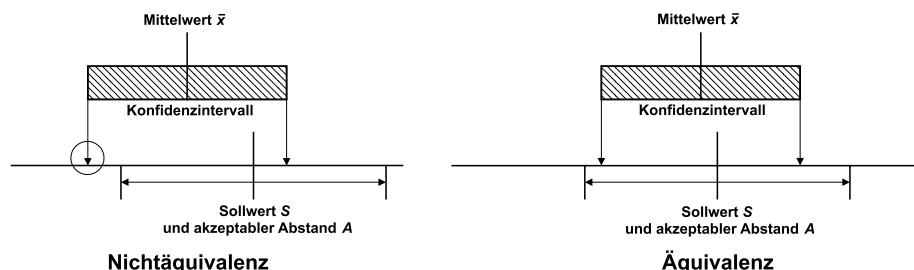


Bild A₆ Äquivalenz und Nichtäquivalenz

Die Abschätzung der Äquivalenz wird über zwei Prüfgrößen vorgenommen. Zur Berechnung der beiden Prüfgrößen PG_u und PG_o dienen Gl. A₂₀ und Gl. A₂₁:

$$PG_u = S - \bar{x} - \frac{s \cdot t}{\sqrt{N}} \quad (\text{Gl. A}_{20})$$

$$PG_0 = S - \bar{x} + \frac{s \cdot t}{\sqrt{N}} \quad (\text{Gl. A}_{21})$$

mit:

PG_u Prüfgröße unten

PG_0 Prüfgröße oben

S Sollwert

\bar{x} Mittelwert der Ergebnisse der Stichprobe

s Standardabweichung der Stichprobenergebnisse

t t -Faktor der einseitigen t -Tabelle mit $f = N - 1$ und (empfohlenen) $P = 95\%$

Der Äquivalenztest zeigt auf dem Niveau von $P = 95\%$ keine Unterschiede zwischen beiden Intervallen, wenn

$$-A \leq PG_u \text{ und } PG_0 \leq A$$

ist, wobei A der absolute Akzeptanzwert ist (also nicht in %!).

Beispiel

Ein Labor hat $N = 6$ Proben analysiert. Deren Kenndaten sind:

$$\begin{array}{ll} \text{Mittelwert} & \bar{x} = 98,8 \text{ mg} \\ \text{Standardabweichung} & s = 0,85 \text{ mg} \\ N & = 6 \end{array}$$

Der Sollwert beträgt $S = 100 \text{ mg}$, der akzeptable Abstand wird mit $A = \pm 2 \text{ mg}$ ($\pm 2\%$) festgelegt, der Test soll auf dem Niveau von $P = 95\%$ erfolgen. Der t -Faktor nach der einseitigen t -Tabelle (Anhang Anh₁₃) mit $f = N - 1 = 6 - 1 = 5$ und $P = 95\%$ beträgt 2,015.

Die Prüfgröße nach unten (PG_u) kann mit Gl. A₂₂ berechnet werden:

$$PG_u = S - \bar{x} - \frac{s \cdot t}{\sqrt{N}} = 100 \text{ mg} - 98,8 \text{ mg} - \frac{0,85 \text{ mg} \cdot 2,015}{\sqrt{6}} = \underline{\underline{0,501 \text{ mg}}} \quad (\text{Gl. A}_{22})$$

Die Prüfgröße nach oben (PG_0) kann mit Gl. A₂₃ berechnet werden:

$$PG_0 = S - \bar{x} + \frac{s \cdot t}{\sqrt{N}} = 100 \text{ mg} - 98,8 \text{ mg} + \frac{0,85 \text{ mg} \cdot 2,015}{\sqrt{6}} = \underline{\underline{1,899 \text{ mg}}} \quad (\text{Gl. A}_{23})$$

Die Bedingungen $-A \leq PG_u \text{ und } PG_0 \leq A$ ($-2 \text{ mg} \leq 0,501 \text{ mg} \text{ und } 1,899 \text{ mg} \leq 2 \text{ mg}$) werden eingehalten, somit sind der Sollwert S und der Mittelwert \bar{x} dem akzeptablen Abstand von 2% äquivalent.

Als Vergleich dazu ist nachfolgend der **Sollwert-t-Test** nach Gl. A₂₄ aufgeführt, bei dem ein akzeptabler Abstand A keine Rolle spielt.

$$PG = \frac{|\bar{x} - S|}{s} \cdot \sqrt{N} = \frac{98,8 \text{ mg} - 100 \text{ mg}}{0,85 \text{ mg}} \cdot \sqrt{6} = \underline{\underline{3,458}} \quad (\text{Gl. A}_{24})$$

Der Schrankenwert mit $P = 95\%$ und $f = 6 - 1 = 5$ beträgt 2,571 (zweiseitig), damit gäbe es auf diesem Signifikanzniveau einen Unterschied zwischen dem Sollwert S und dem Mittelwert \bar{x} .

2. Äquivalenztest: Unterschied zweier Mittelwerte

Die Unterschiede zwischen Mittelwerten zweier Stichprobenreihen sind z.B. zur Beurteilung der Richtigkeit eines neuen analytischen Verfahrens vergleichbar, wenn die Ergebnisse (Mittelwerte) des neuen Verfahrens mit denen eines bereits validierten Verfahrens verglichen werden.

Zur Durchführung des Äquivalenztests auf Mittelwertunterschiede zweier Datenreihen wird zunächst der akzeptable Abstand A beider Mittelwerte festgelegt. Meistens beträgt dieser akzeptable Abstand 1...3%. Dann wird aus den Varianzen der beiden Datenreihen mit einem **einfachen F-Test** geprüft, ob keine **Varianzeninhomogenität** nachgewiesen werden kann (empfohlen wird das Signifikanzniveau $P = 99\%$). Sind die Varianzen *nicht* homogen, ist der Äquivalenztest nicht sinnvoll. Sind die Varianzen homogen, kann nach Gl. A₂₅ eine gemeinsame (gepoolte) Standardabweichung s_d berechnet werden:

$$s_d = \sqrt{\frac{(N_1 - 1) \cdot s_1^2 + (N_2 - 1) \cdot s_2^2}{N_1 + N_2 - 2}} \quad (\text{Gl. A}_{25})$$

Danach werden die beiden Prüfgrößen PG_1 und PG_2 mit Gl. A₂₆ und Gl. A₂₇ berechnet:

$$PG_1 = \frac{|\bar{x}_1 - \bar{x}_2| + A}{s_d} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \quad (\text{Gl. A}_{26})$$

$$PG_2 = \frac{A - |\bar{x}_1 - \bar{x}_2|}{s_d} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} \quad (\text{Gl. A}_{27})$$

mit:

PG_1 Prüfgröße 1

PG_2 Prüfgröße 2

s_d gemeinsame Standardabweichung

A akzeptabler Abstand

\bar{x}_1 Mittelwert 1

\bar{x}_2 Mittelwert 2

N_1 Anzahl der Daten in Reihe 1

N_2 Anzahl der Daten in Reihe 2

Dann werden die beiden Prüfgrößen PG_1 und PG_2 mit den Tabellenwerten der **einseitigen t-Tabelle** mit $f = f_1 + f_2$ verglichen (vorgeschlagen wird $P = 95\%$). Sind **beide** Prüfgrößen PG_1 bzw. PG_2 **größer** als der Tabellenwert, kann auf akzeptable Äquivalenz der Mittelwerte geschlossen werden.

Beispiel

Ein *Validierlabor* hat aus $N = 6$ Mehrfachbestimmungen einer Probe Analysenergebnisse erzielt, deren Kenndaten sind:

Mittelwert	$\bar{x}_V = 100,9 \text{ mg}$
Standardabweichung	$s_V = 0,85 \text{ mg}$
N	= 6

Ein *Routinelabor* hat aus $N = 6$ Mehrfachbestimmungen mit identischer Arbeitsvorschrift der gleichen Probe folgende Werte erzielt:

$$\begin{array}{ll} \text{Mittelwert} & \bar{x}_R = 102,4 \text{ mg} \\ \text{Standardabweichung} & s_R = 0,95 \text{ mg} \\ N & = 6 \end{array}$$

Der akzeptable Abstand A der Mittelwerte wird mit 2,5 mg (ca. 2,5%) festgelegt, die Untersuchung soll auf dem Niveau von $P = 95\%$ erfolgen. Als Erstes wird mit einem F -Test überprüft, ob beide Varianzen homogen sind (Gl. A₂₈):

$$PG = \left(\frac{0,95 \text{ mg}}{0,85 \text{ mg}} \right)^2 = 1,249 \quad (\text{Gl. A}_{28})$$

Der Schrankenwert beträgt 5,050 (aus der F -Tabelle (Anhang) mit $f_1 = f_2 = N - 1 = 6 - 1 = 5$ und $P = 95\%$). Die Varianzen sind homogen, da die PG kleiner als der Schrankenwert ist. Die gemeinsame Standardabweichung s_d nach Gl. A₂₉ beträgt 0,901 mg.

$$s_d = \sqrt{\frac{(6-1) \cdot 0,85^2 \text{ mg}^2 + (6-1) \cdot 0,95^2 \text{ mg}^2}{6+6-2}} = 0,901 \text{ mg} \quad (\text{Gl. A}_{29})$$

Die Prüfgröße PG_1 wird berechnet mit Gl. A₃₀:

$$PG_1 = \frac{|\bar{x}_V - \bar{x}_R| + A}{s_d} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} = \frac{|100,9 \text{ mg} - 102,4 \text{ mg}| + 2,5 \text{ mg}}{0,901 \text{ mg}} \cdot \sqrt{\frac{6 \cdot 6}{6+6}} = 7,689 \quad (\text{Gl. A}_{30})$$

Die Prüfgröße PG_2 wird berechnet mit Gl. A₃₁:

$$PG_2 = \frac{A - |\bar{x}_V - \bar{x}_R|}{s_d} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} = \frac{2,5 \text{ mg} - |100,9 \text{ mg} - 102,4 \text{ mg}|}{0,901 \text{ mg}} \cdot \sqrt{\frac{6 \cdot 6}{6+6}} = 1,922 \quad (\text{Gl. A}_{31})$$

Der Tabellenwert der t -Tabelle (einseitig) mit $f_1 + f_2 = 10$ und $P = 95\%$ beträgt 1,812. Da beide Prüfgrößen **größer** sind als der Schrankenwert der **einseitigen** t -Tabelle, kann auf die Äquivalenz der Mittelwerte hinsichtlich des akzeptablen Abstandes von $A = 2,5$ mg geschlossen werden.

Würde nur ein Mittelwert- t -Test mit den Daten beider Datenreihen durchgeführt, bei dem der akzeptable Abstand keine Rolle spielt, erhält man nach Gl. A₃₂ die Prüfgröße $PG = 2,882$.

$$PG = \frac{|\bar{x}_V - \bar{x}_R|}{s_d} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}} = \frac{|100,9 \text{ mg} - 102,4 \text{ mg}|}{0,901 \text{ mg}} \cdot \sqrt{\frac{6 \cdot 6}{6+6}} = 2,882 \quad (\text{Gl. A}_{32})$$

Der Schrankenwert wäre mit $f=10$ und $P = 95\%$ (zweiseitig) 2,228, die Äquivalenz der beiden Mittelwerte würde auf dem Signifikanzniveau von $P = 95\%$ abgelehnt.

Anwendbarkeit

Häufig ist es aus wirtschaftlichen Gesichtspunkten sinnvoll, den **Validierungsparameter Robustheit** in verschiedenen zeitlichen Stufen zu überprüfen. Die Prüfung der Anwendbarkeit einer Methode erfolgt zu einem Zeitpunkt, nachdem die Robustheit der Methode bereits akzeptiert wurde, z.B. durch Variation der Temperatur oder des pH-Wertes (Methodenrobustheit). Die Anwendbarkeit sollte vor dem Einsatz in der Routineanalytik überprüft werden. Hierbei müssen 2 Fälle unterschieden werden:

- Die Methode wird zukünftig nur im eigenen Labor eingesetzt.
- Die Methode soll auch außerhalb des eigenen Labors zum Einsatz kommen.

Je nach beabsichtigtem Einsatz ist der Aufwand zur Prüfung der Anwendbarkeit unterschiedlich groß. Soll die Methode nur im eigenen Labor zum Einsatz kommen, ist die **Laborpräzision** zu überprüfen. Der Umfang hängt dabei von der konkreten Laborsituation ab. Ist hingegen vorgesehen, dass die Methode zukünftig außerhalb des eigenen Labors eingesetzt wird, sollte die Prüfung der **Vergleichspräzision** erfolgen. Wenn mehrere Labors beteiligt sind, werden meist **Ringversuche** durchgeführt. Diese besitzen besonders im behördlichen Umfeld oder in Auftraglabors eine große Bedeutung.

Äquidistanz

Beim Kalibrieren herrschen dann äquidistante Konzentrationsverhältnisse, wenn die für eine Linearitätsuntersuchung hergestellten Kalibrierlösungen untereinander etwa den gleichen Konzentrationsabstand aufweisen. Eine solche Äquidistanz ist beim **Kalibrieren** anzustreben, da die gewonnenen statistischen Kenndaten (z.B. die **Reststandardabweichung s_y** und die **Verfahrensstandardabweichung V_{x0}**) unter dieser Bedingung die höchste Aussagekraft haben.

Arbeitsbereich

Der Arbeitsbereich ist vom angewandten **Analysenverfahren** abhängig. Er gibt den Konzentrationsbereich an, für den die Parameter **Richtigkeit**, **Präzision** und **Linearität** überprüft wurden und Gültigkeit besitzen. Der Arbeitsbereich wird meistens durch eine **Kalibrierfunktion** beschrieben. Eine **Validierung** umfasst die Gültigkeitsprüfung des Arbeitsbereiches. Für eine Quantifizierung wird der Arbeitsbereich von einer oberen und einer unteren Konzentrationsgrenze eingeschlossen, wobei nicht zwangsläufig der gesamte für die Kalibrierung nützliche Bereich als Arbeitsbereich angegeben werden muss.

Die untere Konzentrationsgrenze kann über die Bestimmungsgrenze definiert werden, also die kleinste zu quantifizierende Konzentration. Unter dieser Grenze ist die Ergebnisunsicherheit für eine quantitative Aussage zu groß. Die obere Konzentrationsgrenze hängt davon ab, ob eine durch **Regression** gefundene **Kalibrierfunktion** und die oben genannten Parameter noch gültig sind. Diese Konzentrationsgrenzen können durch eine Untersuchung auf **Varianzenhomogenität** überprüft werden.

Die **ICH** beispielsweise gibt Empfehlungen für den Arbeitsbereich an. Für die Quantifizierung eines Wirkstoffes z.B. wird der Arbeitsbereich mit der Angabe $\pm 20\%$ um die **Arbeitsbereichsmittel** vorgeschlagen, ausgehend davon, dass die Arbeitsbereichsmittel die Zielkonzentration repräsentiert.

Arbeitsbereichsmittel \bar{x} bzw. \bar{y}

Um die Arbeitsbereichsmittel der x -Werte \bar{x} berechnen zu können, werden die jeweiligen x -Werte (x_i) summiert und die Summe durch die Anzahl der Kalibrierniveaus N geteilt. Die Arbeitsbereichsmittel \bar{x} erhält man aus Gl. A₃₃:

$$\bar{x} = \frac{\sum x_i}{N} \quad (\text{Gl. A}_{33})$$

Analog erhält man die Arbeitsbereichsmitte für die y -Werte \bar{y} nach Gl. A₃₄:

$$\bar{y} = \frac{\sum y_i}{N} \quad (\text{Gl. A}_{34})$$

Die Arbeitsbereichsmitten \bar{x} und \bar{y} werden zur Berechnung von Kenndaten bei einer [Regression](#) zur Ermittlung der Regressionsfunktion benötigt.

Beispiel

Bei einer fotometrischen Messung wurden die in Tabelle A₅ aufgeführten Werte aufgenommen.

Tabelle A₅ Ergebnisse der fotometrischen Messung

x-Werte	y-Werte
Konzentration	Extinktion
in mg/100 mL	
3	0,235
5	0,405
7	0,535
9	0,620
11	0,720
13	0,810
15	0,850

Die Anwendung von Gl. A₃₃ und Gl. A₃₄ ergibt für die Arbeitsbereichsmitten die Werte von Gl. A₃₅ und Gl. A₃₆:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{3 + 5 + 7 + 9 + 11 + 13 + 15}{7} = 9 \text{ mg/100mL} \quad (\text{Gl. A}_{35})$$

$$\bar{y} = \frac{\sum y_i}{N} = \frac{0,235 + 0,405 + 0,535 + 0,620 + 0,720 + 0,810 + 0,850}{7} = 0,596 \quad (\text{Gl. A}_{36})$$

Arzneibuch

Das Arzneibuch (Pharmakopöe) ist eine amtliche Sammlung anerkannter Werke über die Qualität, Analytik, Lagerung und Bezeichnung pharmazeutischer Wirkstoffe und der daraus hergestellten [Arzneimittel](#). Das Arzneibuch in Deutschland hat als gesetzliche Grundlage den § 55 des Arzneimittelgesetzes (AMG): «Das Arzneibuch hat die Aufgabe, die Volksgesundheit mit Hilfe anerkannter, gemeinsamer Regeln zu fördern, die von den Verantwortlichen im Gesundheitswesen und allen Personen, die sich mit der Qualität der Arzneimittel befassen, zu beachten und einzuhalten sind.» [A2]

Das Deutsche Arzneibuch (DAB) ist in mehr als 10 Ausgaben erschienen. Die wichtigsten Ausgaben sind:

- DAB 6, 1926
- DAB 7, 1968
- DAB 8, 1978